On almost randomizing channels with a short Kraus decomposition

Guillaume Aubrun

ON ALMOST RANDOMIZING CHANNELS WITH A SHORT KRAUS DECOMPOSITION

GUILLAUME AUBRUN

Abstract. For large d, we study quantum channels on $\mathbf{C}^d$ obtained by selecting randomly N independent Kraus operators according to a probability measure μ on the unitary group $\mathcal{U}(d)$. When μ is the Haar measure, we show that for $N \succcurlyeq d/\varepsilon^2$, such a channel is ε-randomizing with high probability, which means that it maps every state within distance ε/d (in operator norm) of the maximally mixed state. This slightly improves on a result by Hayden, Leung, Shor and Winter by optimizing their discretization argument. Moreover, for general μ, we obtain a ε-randomizing channel provided $N \succcurlyeq d(\log d)^6/\varepsilon^2$. For $d = 2^k$ (k qubits), this includes Kraus operators obtained by tensoring k random Pauli matrices. The proof uses recent results on empirical processes in Banach spaces.

1. Introduction

The completely randomizing quantum channel on $\mathbf{C}^d$ maps every state to the maximally mixed state ρ_*. This channel is used to construct perfect encryption systems (see [1] for formal definitions). However it is a complex object in the following sense: any Kraus decomposition must involve at least d^2 operators. It has been shown by Hayden, Leung, Shor and Winter [12] that this "ideal" channel can be efficiently emulated by lower-complexity channels, leading to approximate encryption systems. The key point is the existence of good approximations with much shorter Kraus decompositions. More precisely, say that a quantum channel Φ on $\mathbf{C}^d$ is ε-randomizing if for any state ρ, $\|\Phi(\rho) - \rho_*\|_\infty \leqslant \varepsilon/d$. The existence of ε-randomizing channels with $o(d^2)$ Kraus operators has several other implications [12], such as counterexamples to multiplicativity conjectures [17].

It has been proved in [12] that if (U_i) denote independent random matrices Haar-distributed on the unitary group $\mathcal{U}(d)$, then the quantum channel

$$\Phi : \rho \mapsto \frac{1}{N} \sum_{j=1}^{N} U_i \rho U_i^\dagger \tag{1}$$

is ε-randomizing with high probability provided $N \geqslant Cd \log d/\varepsilon^2$ for some constant C. The proof uses a discretization argument and the fact that the Haar measure satisfies subgaussian estimates. We show a simple trick that allows to drop a $\log d$ factor: Φ is ε-randomizing when $N \geqslant Cd/\varepsilon^2$, this is our theorem 1.

The Haar measure is a nice object from the theoretical point of view, but is often too complicated to implement for concrete situations. Let us say that a measure μ on $\mathcal{U}(d)$ is isotropic when $\int U\rho U^\dagger d\mu(U) = \rho_*$ for any state ρ. When $d = 2^k$, an example of isotropic measure is given by assigning equal masses at k-wise tensor products of Pauli operators.

The following question was asked in [12]: is the quantum channel Φ defined as (1) ε-randomizing when (U_i) are distributed according to any isotropic probability measure on $\mathcal{U}(d)$? We answer positively this question when $N \geqslant Cd \log^6 d/\varepsilon^2$. This is our main result and appears as theorem 2. Note that for non-Haar measures, previous results appearing in the literature [12, 2, 8] involved the weaker trace-norm approximation $\|\Phi(\rho) - \rho_*\|_1 \leqslant \varepsilon$.

As opposed to the Haar measure, the measure μ need not have subgaussian tails, and we need more sophisticated tools to prove theorem 2. We use recent results on suprema of empirical processes in Banach spaces. After early work by Rudelson [15] and Guédon–Rudelson [11], a general sharp inequality was obtained by Guédon, Mendelson, Pajor and Tomczak–Jaegermann [10]. This inequality is valid in any Banach space with a sufficiently regular equivalent norm, such as ℓ_1^d. The problem of ε-randomizing channels involves the supremum of an empirical process in the trace-class space S_1^d (non-commutative analogue of ℓ_1^d), which enters perfectly this setting.

The paper is organized as follows. Section 2 contains background and precise statements of the theorems. Theorem 1 (for Haar measure) is proved in section 3. Theorem 2 (for a general measure) is proved in section 4. An appendix contains the needed facts about geometry and probability in Banach spaces.

Acknowledgement. I thank Andreas Winter for several e-mail exchanges on the topic, and I am very grateful to Alain Pajor for showing me that the results of [10] can be applied here.

2. Background and presentation of results

Thoughout the paper, the letter C and c denote absolute constants whose value may change from occurrence to occurrence. We usually do not pay too much attention to the value of these constants.

2.1. **Schatten classes.** We write $\mathcal{M}(\mathbf{C}^d)$ for the space of complex $d \times d$ matrices. If $A \in \mathcal{M}(\mathbf{C}^d)$, let $s_1(A), \ldots, s_d(A)$ denote the *singular values* of A (defined as the square roots of the eigenvalues of $AA^\dagger$). For $1 \leqslant p \leqslant \infty$, the *Schatten p-norm* is defined as

$$\|A\|_p = \left(\sum_{i=1}^d s_i(A)^p \right)^{1/p}.$$

For $p = \infty$, the definition should be understood as $\|A\|_\infty = \max s_i(A)$ and coincides with the usual operator norm. It is well-known (see [5], section IV.2) that $(\mathcal{M}(\mathbf{C}^d), \|\cdot\|_p)$ is a complex normed space, denoted S_p^d and called *Schatten class*. The space S_p^d is the non-commutative analogue of the space ℓ_p^d. We write $B(S_p^d)$ for the unit ball of S_p^d.

The Schatten 2-norm (sometimes called Hilbert–Schmidt or Frobenius norm) is a Hilbert space norm associated to the inner product $\langle A, B \rangle = \operatorname{Tr} A^\dagger B$. This Hermitian structure allows to identify $\mathcal{M}(\mathbf{C}^d)$ with its dual space. Duality on Schatten norms holds as in the commutative case: if p and q are conjugate exponents (i.e. $1/p + 1/q = 1$), then the normed space dual to S_p^d coincides with S_q^d.

2.2. **Completely positive maps.** We write $\mathcal{M}_{\mathrm{sa}}(\mathbf{C}^d)$ (resp. $\mathcal{M}_+(\mathbf{C}^d)$) for the set of self-adjoint (resp. positive semi-definite) $d \times d$ matrices. A linear map $\Phi : \mathcal{M}(\mathbf{C}^d) \to \mathcal{M}(\mathbf{C}^d)$ is said to *preserve positivity* if $\Phi(\mathcal{M}_+(\mathbf{C}^d)) \subset \mathcal{M}_+(\mathbf{C}^d)$. Moreover, Φ is said to be *completely positive* if for any $k \in \mathbf{N}$, the map

$$\Phi \otimes \mathrm{Id}_{\mathcal{M}(\mathbf{C}^k)} : \mathcal{M}(\mathbf{C}^d \otimes \mathbf{C}^k) \to \mathcal{M}(\mathbf{C}^d \otimes \mathbf{C}^k)$$

preserves positivity. We use freely the canonical identification $\mathcal{M}(\mathbf{C}^d) \otimes \mathcal{M}(\mathbf{C}^k) \approx \mathcal{M}(\mathbf{C}^d \otimes \mathbf{C}^k)$.

If $(e_i)_{0 \leqslant i \leqslant d-1}$ denotes the canonical basis of $\mathbf{C}^d$, let $E_{ij} = |e_i\rangle\langle e_j|$. To $\Phi : \mathcal{M}(\mathbf{C}^d) \to \mathcal{M}(\mathbf{C}^d)$ we associate $A_\Phi \in \mathcal{M}(\mathbf{C}^d \otimes \mathbf{C}^d)$ defined as

$$A_\Phi = \sum_{i,j=1}^d E_{ij} \otimes \Phi(E_{ij}).$$

The matrix A_Φ is called the *Choi matrix* of Φ ; it is well-known [7] that Φ is completely positive if and only if A_Φ is positive. Therefore, the set of completely positive operators on $\mathcal{M}(\mathbf{C}^d)$ is in one-to-one correspondence with $\mathcal{M}_+(\mathbf{C}^d \otimes \mathbf{C}^d)$. This correspondence is known as the *Choi–Jamiołkowski isomorphism*.

The spectral decomposition of A_Φ implies now the following: any completely positive map Φ on $\mathcal{M}(\mathbf{C}^d)$ can be decomposed as

$$\Phi : X \mapsto \sum_{i=1}^{N} V_i X V_i^\dagger. \tag{2}$$

Here $V_1, \ldots, V_N$ are elements of $\mathcal{M}(\mathbf{C}^d)$. This decomposition is called a *Kraus decomposition* of Φ of length N. The minimal length of a Kraus decomposition of Φ (called *Kraus rank*) is equal to the rank of the Choi matrix A_Φ. In particular it is always bounded by d^2.

2.3. **States and the completely depolarizing channel.** A *state* on $\mathbf{C}^d$ is a element of $\mathcal{M}_+(\mathbf{C}^d)$ with trace 1. We write $\mathcal{D}(\mathbf{C}^d)$ for the set of states ; it is a compact convex set with (real) dimension $d^2 - 1$. If $x \in \mathbf{C}^d$ is a unit vector, we write $P_x = |x\rangle\langle x|$ for the associated rank one projector. The state P_x is called a *pure state*, and it follows from spectral decomposition that any state is a convex combination of pure states. A central role is played by the *maximally mixed state* $\rho_* = \mathrm{Id}/d$ (ρ_* is sometimes called the *random state*).

A *quantum channel* $\Phi : \mathcal{M}(\mathbf{C}^d) \to \mathcal{M}(\mathbf{C}^d)$ is a completely positive map which preserves trace: for any $X \in \mathcal{M}(\mathbf{C}^d), \mathrm{Tr}\,\Phi(X) = \mathrm{Tr}\,X$. Note that a quantum channel maps states to states. The trace-preserving condition reads on the Kraus decomposition (2) as

$$\sum_{i=1}^{N} V_i^\dagger V_i = \mathrm{Id}.$$

An example of quantum channel that plays a central role in quantum information theory is the *(completely) randomizing channel* (also called *completely depolarizing channel*) $R : \mathcal{M}(\mathbf{C}^d) \to \mathcal{M}(\mathbf{C}^d)$.

$$R : X \to \mathrm{Tr}\,X \cdot \frac{\mathrm{Id}}{d}.$$

The randomizing channel maps every state to ρ_*. The Choi matrix of R is $A_R = \frac{1}{d}\mathrm{Id}_{\mathbf{C}^d \otimes \mathbf{C}^d}$. Since A_R has full rank, any Kraus decomposition of R must have length (at least) d^2. An explicit decomposition can be written using Fourier-type unitary operators: let $\omega = \exp(2i\pi/d)$ and A and B the matrices defined as

$$A(e_j) = e_{j+1 \bmod d} \qquad B(e_j) = \omega^j e_j. \tag{3}$$

For $1 \leqslant j, k \leqslant d$, define $V_{j,k}$ as the product $B^j A^k$. Note that $V_{j,k}$ belongs to the unitary group $\mathcal{U}(d)$. A routine calculation (see also section 2.5) shows that for any $X \in \mathcal{M}(\mathbf{C}^d)$,

$$\frac{1}{d^2} \sum_{j,k=1}^{d} V_{j,k} X V_{j,k}^\dagger = \mathrm{Tr}\,X \cdot \frac{\mathrm{Id}}{d}.$$

This is a Kraus decomposition of the randomizing channel.

2.4. **ε-randomizing channels.** We are interested in approximating the randomizing channel R by channels with low Kraus rank. Following Hayden, Leung, Shor and Winter [12], a quantum channel Φ is called *ε-randomizing* if for any state $\rho \in \mathcal{D}(\mathbf{C}^d)$,

$$\|\Phi(\rho) - \rho_*\|_\infty \leqslant \frac{\varepsilon}{d}.$$

It is equivalent to say that the spectrum of $\Phi(\rho)$ is contained in $[(1-\varepsilon)/d, (1+\varepsilon)/d]$ for any state ρ. It has been proved in [12] that there exist ε-randomizing channels with Kraus rank equal to $Cd\log d/\varepsilon^2$ for some constant d. This is much smaller that d^2 (the Kraus rank of R). The construction is simple: generate independent random Kraus operators according to the Haar measure on $\mathcal{U}(d)$ and show that the induced quantum channel is ε-randomizing with nonzero probability. A key step in the proof is a discretization argument. We show that a simple trick improves the efficiency of the argument from [12] to prove the following

Theorem 1 (Haar-generated ε-randomizing channels)**.** *Let $(U_i)_{1\leqslant i\leqslant N}$ be independent random matrices Haar-distributed on the unitary group $\mathcal{U}(d)$. Let $\Phi : \mathbf{C}^d \to \mathbf{C}^d$ be the quantum channel defined by*

$$\Phi(\rho) = \frac{1}{N}\sum_{i=1}^N U_i \rho U_i^\dagger.$$

Assume that $0 < \varepsilon < 1$ and $N \geqslant Cd/\varepsilon^2$. Then the channel Φ is ε-randomizing with nonzero probability.

As often with random constructions, we actually prove that the conclusion holds true with *large* probability: the probability of failure is exponentially small in d.

It is clear that the way N depends on d is optimal: if Φ is a ε-randomizing channel with $\varepsilon < 1$, its Kraus rank must be at least d. This is because for any pure state P_x, $\Phi(P_x)$ must have full rank. The dependence in ε is sharp for channels as constructed here, since lemma 2 below is sharp. However, it is not clear whether families of ε-randomizing channels with a better dependence in ε can be found using a different construction, possibly partially deterministic.

One checks (using the value $c = 1/6$ from [12] in lemma 3 and optimizing over the net size) that the constant in theorem 1 can the chosen to, e.g., $C = 150$. This is presumably far from optimal.

2.5. **Isotropic measures on unitary matrices.** Although the quantum channels constructed in theorem 1 have minimal Kraus rank, it can be argued that Haar-distributed random matrices are hard to generate in real-life situations. We introduce a wide class of measures on $\mathcal{U}(d)$ that may replace the Haar measure.

Definition. *Say that a probability measure μ on $\mathcal{U}(d)$ is* isotropic *if for any $X \in \mathcal{M}(\mathbf{C}^d)$,*

$$\int_{\mathcal{U}(d)} UXU^\dagger d\mu(U) = \operatorname{Tr} X \cdot \frac{\mathrm{Id}}{d}.$$

Similarly, a $\mathcal{U}(d)$-valued random vector is called isotropic if its law is isotropic.

Lemma 1. *Let $U = (U_{ij})$ be a $\mathcal{U}(d)$-valued random vector. The following assertions are equivalent*

1. *U is isotropic.*
2. *For any $X \in \mathcal{M}(\mathbf{C}^d)$, $\mathbf{E}|\operatorname{Tr} UX^\dagger|^2 = \frac{1}{d}\|X\|_2^2$.*
3. *For any indices i, j, k, l, $\mathbf{E}U_{ij}\overline{U_{kl}} = \frac{1}{d}\delta_{i,k}\delta_{j,l}$.*

Proof. Implications (3) $\Rightarrow$ (1) and (3) $\Rightarrow$ (2) are easily checked by expansion. For (1) $\Rightarrow$ (3), simply take $X = |e_j\rangle\langle e_k|$. Identity (2) implies after polarization that for any $A, B \in \mathcal{M}(\mathbf{C}^d)$,

$$\mathbf{E}\left[\overline{\operatorname{Tr}(UA^\dagger)}\operatorname{Tr}(UB^\dagger)\right] = \frac{1}{d}\operatorname{Tr}(AB^\dagger),$$

from which (3) follows. □

Condition (3) of the lemma means that the covariance matrix of U — which is an element of $\mathcal{M}(\mathcal{M}(\mathbf{C}^d))$ — is a multiple of the identity matrix.

Of course the Haar measure is isotropic. Other examples are provided by discrete measures. Let $\mathscr{U} = \{U_1, \dots, U_{d^2}\}$ be a family of unitary matrices, which are mutually orthogonal in the following sense: if $i \neq j$, then $\operatorname{Tr} U_i^\dagger U_j = 0$. For example, one can take $\mathscr{U} = \{B^j A^k\}_{1 \leqslant j,k \leqslant d}$, A, B defined as (3). Then the uniform probability measure on $\mathscr{U}$ is isotropic. Indeed, any $X \in \mathcal{M}(\mathbf{C}^d)$ can be decomposed as $X = \sum x_i U_i$ and condition (2) of lemma 1 is easily checked.

If we specialize to $d = 2$, we obtain a random Pauli operator: assign probability $1/4$ to each of the following matrices to get a isotropic measure

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is straightforward to check that isotropic vectors tensorize: if $X_1 \in \mathcal{U}(d_1)$ and $X_2 \in \mathcal{U}(d_2)$ are isotropic, so is $X_1 \otimes X_2 \in \mathcal{U}(d_1 d_2)$. If we work on $\mathcal{M}((\mathbf{C}^2)^{\otimes k})$, which corresponds to a set of k qubits, a natural isotropic measure is therefore obtained by choosing independently a Pauli matrix on each qubit, i.e. assigning mass $1/4^k$ to the matrix $\sigma_{i_1} \otimes \cdots \otimes \sigma_{i_k}$ for any $i_1, \dots, i_k \in \{0, 1, 2, 3\}^k$.

2.6. **ε-randomizing channels for an isotropic measure.** We can now state our main theorem asserting that up to logarithmic terms, the Haar measure can be replaced in theorem 1 by simpler notions of randomness. We first state our result

Theorem 2 (General ε-randomizing channels)**.** *Let μ be an isotropic measure on the unitary group $\mathcal{U}(d)$. Let $(U_i)_{1 \leqslant i \leqslant N}$ be independent μ-distributed random matrices, and $\Phi : \mathbf{C}^d \to \mathbf{C}^d$ be the quantum channel defined as*

$$\Phi(\rho) = \frac{1}{N} \sum_{i=1}^{N} U_i \rho U_i^\dagger. \tag{4}$$

Assume that $0 < \varepsilon < 1$ and $N \geqslant Cd(\log d)^6/\varepsilon^2$. Then the channel Φ is ε-randomizing with nonzero probability.

Theorem 2 applies in particular for product of random Pauli matrices as described in the previous section. It is of interest for certain cryptographic applications to know that ε-randomizing channels can be realized using Pauli matrices.

As opposed to theorem 1, the conclusion of theorem 2 is not proved to hold with exponentially large probability. Applying the theorem with $\varepsilon\eta$ instead of ε and using Markov inequality shows that Φ is ε-randomizing with probability larger than $1 - \eta$ provided $N \geqslant Cd \log^6 d/(\varepsilon^2\eta^2)$.

Theorem 2 could be quickly deduced from a theorem appearing in [10]. However, the proof of [10] is rather intricate and uses Talagrand's majorizing measures in a central way. We give here a proof of our theorem which uses the simpler Dudley integral instead, giving the same result. We however rely an a entropy lemma from [10], which appears as lemma A5 in the appendix.

The $\log^6 d$ appearing in theorem 2 is certainly non optimal (see remarks at the end of the paper). However, some power of $\log d$ is needed, as shown by the next proposition.

Proposition. *Let A, B defined as (3) and μ be the uniform measure on the set $\{B^j A^k\}_{1 \leqslant j,k \leqslant d}$. Consider (X_i) independent μ-distributed random unitary matrices. If the quantum channel Φ defined as (4) is $\frac{1}{2}$-randomizing with probability larger than $1/2$, then $N \geqslant cd \log d$.*

Proof. We will rely on the following standard result in elementary probability theory known as the coupon collector's problem (see [9], Chapter 1, example 5.10): if we choose independently and uniformly random elements among a set of d elements, the mean (and also the median) number of choices before getting all elements at least once is equivalent to $d \log d$ for large d.

In our case, remember that $\omega = \exp(2i\pi/d)$ and for $0 \leqslant j \leqslant d-1$, define $x_j \in \mathbf{C}^d$ as

$$x_j = \left(\frac{1}{\sqrt{d}}, \frac{\omega^j}{\sqrt{d}}, \frac{\omega^{2j}}{\sqrt{d}}, \ldots, \frac{\omega^{(d-1)j}}{\sqrt{d}} \right).$$

Note that $\mathscr{B} = (x_j)_{0 \leqslant j \leqslant d-1}$ is an orthonormal basis of $\mathbf{C}^d$ and that $B^j A^k x_0 = x_j$. Consequently, if U is μ-distributed, the random state $U P_{x_0} U^\dagger$ equals P_{x_j} with probability $1/d$. In the basis $\mathscr{B}$, the matrix $\Phi(P_{x_0})$ is diagonal. Note that if Φ is $\frac{1}{2}$-randomizing, then $\Phi(P_{x_0})$ must have full rank. The reduction to the coupon collector's problem is now immediate. □

3. Proof of theorem 1: Haar-distributed unitary operators.

The scheme of the proof is similar to [12]. We need two lemmas from there. The first is a deviation inequality sometimes known as Bernstein's inequality. The second is proved by a volumetric argument.

Lemma 2 (Lemma II.3 in [12])**.** *Let φ, ψ be pure states on $\mathbf{C}^d$ and $(U_i)_{1 \leqslant i \leqslant N}$ be independent Haar-distributed random unitary matrices. Then for every $0 < \delta < 1$,*

$$\mathbf{P}\left(\left| \frac{1}{N} \sum_{i=1}^N \mathrm{Tr}(U_i \varphi U_i^\dagger \psi) - \frac{1}{d} \right| \geqslant \frac{\delta}{d} \right) \leqslant 2 \exp(-c\delta^2 N)$$

Lemma 3 (Lemma II.4 in [12])**.** *For $0 < \delta < 1$ there exists a set $\mathcal{N}$ of pure states on $\mathbf{C}^d$ with $|\mathcal{N}| \leqslant (5/\delta)^{2d}$, such that for every pure state φ on $\mathbf{C}^d$, there exists $\varphi_0 \in \mathcal{N}$ such that $\|\varphi - \varphi_0\|_1 \leqslant \delta$. Such a set $\mathcal{N}$ is called a δ-net.*

The improvement on the result of [12] will follow from the next lemma

Lemma 4 (Computing norms on nets)**.** *Let $\Delta : \mathcal{B}(\mathbf{C}^d) \to \mathcal{B}(\mathbf{C}^d)$ be a Hermitian-preserving linear map. Let A be the quantity*

$$A = \sup_{\varphi \in \mathcal{D}(\mathbf{C}^d)} \|\Delta(\varphi)\|_\infty = \sup_{\varphi, \psi \in \mathcal{D}(\mathbf{C}^d)} |\mathrm{Tr}\, \psi \Delta(\varphi)|$$

Let $0 < \delta < 1/2$ and $\mathcal{N}$ be a δ-net as provided by lemma 3. We can evaluate A as follows

$$A \leqslant \frac{1}{1-2\delta} B,$$

where

$$B = \sup_{\varphi_0, \psi_0 \in \mathcal{N}} |\mathrm{Tr}\, \psi_0 \Delta(\varphi_0)|$$

Proof of lemma 4. First note that for any self-adjoint operators $a, b \in \mathcal{B}(\mathbf{C}^d)$, we have

$$|\mathrm{Tr}\, b \Delta(a)| \leqslant A \|a\|_1 \|b\|_1. \tag{5}$$

By a convexity argument, the supremum in A can be restricted to pure states. Given pure states $\varphi, \psi \in \mathcal{D}(\mathbf{C}^d)$, let $\varphi_0, \psi_0 \in \mathcal{N}$ so that $\|\varphi - \varphi_0\|_1 \leqslant \delta$, , $\|\psi - \psi_0\|_1 \leqslant \delta$. Then

$$|\mathrm{Tr}\, \psi \Delta(\varphi)| \leqslant |\mathrm{Tr}(\psi - \psi_0)\Delta(\varphi)| + |\mathrm{Tr}\, \psi_0 \Delta(\varphi - \varphi_0)| + |\mathrm{Tr}\, \psi_0 \Delta(\varphi_0)|$$

Using twice (5) and taking supremum over φ, ψ gives $A \leqslant \delta A + \delta A + B$, hence the result. □

Proof of the theorem. Let R be the randomizing channel. Fix a $\frac{1}{4}$-net $\mathcal{N}$ with $|\mathcal{N}| \leqslant 20^{2d}$, as provided by lemma 3. Let $\Delta = R - \Phi$ and A, B as in lemma 4. Here A and B are random quantities and it follows from lemma 4 that

$$\mathbf{P}\left(A \geqslant \frac{\varepsilon}{d}\right) \leqslant \mathbf{P}\left(B \geqslant \frac{\varepsilon}{2d}\right).$$

Using the union bound and lemma 2, we get

$$\mathbf{P}\left(B \geqslant \frac{\varepsilon}{2d}\right) \leqslant 20^{4d} \cdot 2\exp(-c\varepsilon^2 N/4).$$

This is less that 1 if $N \geqslant Cd/\varepsilon^2$, for some constant C. □

4. Proof of theorem 2: general unitary operators.

A Bernoulli random variable is a random variable ε so that $\mathbf{P}(\varepsilon = 1) = \mathbf{P}(\varepsilon = -1) = 1/2$. Recall that C denotes an absolute constant whose value may change from occurrence to occurrence. We will derive theorem 2 from the following lemma.

Lemma 5. *Let $U_1, \dots, U_N \in \mathcal{U}(d)$ be* deterministic *unitary operators and let (ε_i) be a sequence of independent Bernoulli random variables. Then*

$$\mathbf{E}_\varepsilon \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \sum_{i=1}^N \varepsilon_i U_i \rho U_i^\dagger \right\|_\infty \leqslant C(\log d)^{5/2} \sqrt{\log N} \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \sum_{i=1}^N U_i \rho U_i^\dagger \right\|_\infty^{1/2}. \tag{6}$$

Proof of theorem 2 (assuming lemma 5). Let μ be an isotropic measure on $\mathcal{U}(d)$ and (U_i) be independent μ-distributed random unitary matrices. Let M be the random quantity

$$M = \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \frac{1}{N} \sum_{i=1}^N U_i \rho U_i^\dagger - \frac{\mathrm{Id}}{d} \right\|_\infty$$

We are going to show that $\mathbf{E}M$ is small. The first step is a standard symmetrization argument. Let (U_i') be independent copies of (U_i) and (ε_i) be a sequence of independent Bernoulli random variables. We explicit as a subscript the random variables with repsect to which expectation is taken

$$\begin{aligned}
\mathbf{E}M &\leqslant \mathbf{E}_{U,U'} \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \frac{1}{N} \sum_{i=1}^N U_i \rho U_i^\dagger - U_i' \rho U_i'^\dagger \right\|_\infty \\
&= \mathbf{E}_{U,U',\varepsilon} \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \frac{1}{N} \sum_{i=1}^N \varepsilon_i (U_i \rho U_i^\dagger - U_i' \rho U_i'^\dagger) \right\|_\infty \\
&\leqslant 2\mathbf{E}_{U,\varepsilon} \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \frac{1}{N} \sum_{i=1}^N \varepsilon_i U_i \rho U_i^\dagger \right\|_\infty
\end{aligned}$$

The inequality of the first line is Jensen's inequality for $\mathbf{E}_{U'}$, while the equality on the second line holds since the distribution of $\rho \mapsto U_i \rho U_i^\dagger - U_i' \rho U_i'^\dagger$ is symmetric (as a $\mathcal{M}(\mathcal{M}(\mathbf{C}^d), \mathcal{M}(\mathbf{C}^d))$-valued

random vector). We then decouple the expectations using lemma 5 for fixed (U_i).

$$\begin{aligned}\mathbf{E}M &\leqslant \frac{C}{\sqrt{N}}(\log d)^{5/2}\sqrt{\log N}\mathbf{E}\sup_{\rho\in\mathcal{D}(\mathbf{C}^d)}\left\|\frac{1}{N}\sum_{i=1}^N U_i\rho U_i^\dagger\right\|_\infty^{1/2}\\ &\leqslant \frac{C}{\sqrt{N}}(\log d)^{5/2}\sqrt{\log N}\mathbf{E}\sqrt{M+\frac{1}{d}}\\ &\leqslant \frac{C}{\sqrt{N}}(\log d)^{5/2}\sqrt{\log N}\sqrt{\mathbf{E}M+\frac{1}{d}}\end{aligned}$$

Using the elementary implication

$$X\leqslant\alpha\sqrt{X+\beta}\Longrightarrow X\leqslant\alpha^2+\alpha\sqrt{\beta}$$

we find that $\mathbf{E}M\leqslant\varepsilon/d$ provided $N\geqslant Cd\log^6 d/\varepsilon^2$. □

It remains to prove lemma 5. We will use several standard concepts from geometry and probability in Banach spaces. All the relevant definitions and statements are postponed to the next section.

Proof of lemma 5. Let Z be the quantity appearing in the left-hand side of (6). By a convexity argument, the supremum is attained for an extremal ρ, i.e. a pure state $P_x=|x\rangle\langle x|$ for some unit vector x. Since the operator norm itself can be written as a supremum over unit vectors, we get

$$Z=\sup_{|x|=|y|=1}\left|\sum_{i=1}^N\varepsilon_i|\langle y|U_i|x\rangle|^2\right|=\sup_{|x|=|y|=1}\left|\sum_{i=1}^N\varepsilon_i|\operatorname{Tr}U_i|x\rangle\langle y||^2\right|\leqslant\sup_{A\in B(S_1^d)}\left|\sum_{i=1}^N\varepsilon_i|\operatorname{Tr}U_iA|^2\right|.$$

The last inequality follows from the fact that $B(S_1^d)=\operatorname{conv}\{|x\rangle\langle y|,|x|=|y|=1\}$. Let $\Phi:B(S_1^d)\to\mathbf{R}^N$ defined as

$$\Phi(A)=(|\operatorname{Tr}U_1A|^2,\dots,|\operatorname{Tr}U_NA|^2).$$

We now apply Dudley's inequality (theorem A2 in the next section) with $K=\Phi(B(S_1^d))$ to estimate $\mathbf{E}Z$ using covering numbers. This yields

$$\mathbf{E}Z\leqslant C\int_0^\infty\sqrt{\log N(\Phi(B(S_1^d)),|\cdot|,\varepsilon)}d\varepsilon$$

where $|\cdot|$ denotes the Euclidean norm on $\mathbf{R}^N$. Define a distance δ on $B(S_1^d)$ as

$$\delta(A,B)=|\Phi(A)-\Phi(B)|=\left(\sum_{i=1}^N\left||\operatorname{Tr}U_iA|^2-|\operatorname{Tr}U_iB|^2\right|^2\right)^{1/2}.$$

We are led to the estimate

$$\mathbf{E}Z\leqslant C\int_0^\infty\sqrt{\log N(B(S_1^d),\delta,\varepsilon)}d\varepsilon.$$

Using the inequality $\left||a|^2-|b|^2\right|\leqslant|a-b|\cdot|a+b|$, the metric δ can be upped bounded as follows

$$\delta(A,B)^2\leqslant\left(\sum_{i=1}^N|\operatorname{Tr}U_i(A+B)|^2\right)\sup_{1\leqslant i\leqslant N}|\operatorname{Tr}U_i(A-B)|^2.$$

Let us introduce a new norm $|||\cdot|||$ on $\mathcal{M}(\mathbf{C}^d)$

$$|||A|||=\sup_{1\leqslant i\leqslant N}|\operatorname{Tr}U_iA|.$$

Let θ be the number equal to

$$\theta := \sup_{A \in B(S_1^d)} \sum_{i=1}^N |\operatorname{Tr} U_i A|^2 = \sup_{\rho \in \mathcal{D}(\mathbf{C}^d)} \left\| \sum_{i=1}^N U_i \rho U_i^\dagger \right\|_\infty .$$

We get that for $A, B \in B(S_1^d)$, $\delta(A,B) \leqslant 2\theta |||A-B|||$, and therefore

$$\mathbf{E}Z \leqslant C\theta \int_0^\infty \sqrt{\log N(B(S_1^d), |||\cdot|||, \varepsilon)} d\varepsilon.$$

It remains to bound this new entropy integral. We split it into three parts, for ε_0 to be determined. If ε is large ($\varepsilon > 1$), since $\|U_i\|_\infty = 1$, we get that $|||\cdot||| \leqslant \|\cdot\|_1$. This means that $N(B(S_1^d), |||\cdot|||, \varepsilon) = 1$ and the integrand is zero. If ε is small ($0 < \varepsilon < \varepsilon_0$), we use the volumetric argument of lemma A1

$$N(B(S_1^d), |||\cdot|||, \varepsilon) \leqslant N(B(S_1^d), \|\cdot\|_1, \varepsilon) \leqslant (3/\varepsilon)^{2d^2}.$$

In the intermediate range ($\varepsilon_0 \leqslant \varepsilon \leqslant 1$), let $q = \log d$ and $p = 1 + 1/(\log d - 1)$ be the conjugate exponent. We are going to approximate the Schatten 1-norm by the Schatten p-norm. It is elementary to check that for $A \in \mathcal{M}(\mathbf{C}^d)$, $\|A\|_q \leqslant e\|A\|_\infty$. By dualizing

$$\|A\|_1 \leqslant e\|A\|_p \Longrightarrow N(B(S_1^d), |||\cdot|||, \varepsilon) \leqslant N(B(S_p^d), |||\cdot|||, \varepsilon/e).$$

We are now in position to apply lemma A5 to the space $E = S_p^d$. By theorems A3 and A4, the 2-convexity constant of S_p^d and the type 2 constant of S_q^d (see next section for definitions) are bounded as follows

$$T_2(S_q^d) \leqslant \lambda(S_p^d) \leqslant \sqrt{q-1} \leqslant \sqrt{\log d}.$$

Since $\|U_i\|_q \leqslant e$, the inequality given by lemma A5 is

$$\sqrt{\log N(B(S_1^d), |||\cdot|||, \varepsilon)} \leqslant \frac{C}{\varepsilon} (\log d)^{3/2} \sqrt{\log N}.$$

We now gather all the estimations

$$\int_0^\infty \sqrt{\log N(B(S_1^d), |||\cdot|||, \varepsilon)} d\varepsilon \leqslant \int_0^{\varepsilon_0} \sqrt{2d^2} \log(3/\varepsilon) d\varepsilon + C(\log d)^{3/2} \sqrt{\log N} \int_{\varepsilon_0}^1 \frac{1}{\varepsilon} d\varepsilon.$$

Choosing $\varepsilon_0 = 1/d$, an immediate computation shows that

$$\int_0^\infty \sqrt{\log N(B(S_1^d), |||\cdot|||, \varepsilon)} d\varepsilon \leqslant C(\log d)^{5/2} \sqrt{\log N}.$$

This concludes the proof of the lemma. □

Appendix : Geometry of Banach spaces

In this last section, we gather several definitions and results from geometry and probability in Banach spaces. We denote by $(E, \|\cdot\|)$ a real or complex Banach space (actually, in our applications E will be finite-dimensional). We denote by $(E^*, \|\cdot\|_*)$ the dual Banach space.

4.1. **Covering numbers.**

Definition. *If (K,δ) is a compact metric space, the* covering number *or* entropy number $N(K,\delta,\varepsilon)$ *is defined to be the smallest cardinality M of a set $\{x_1,\dots,x_M\}\subset K$ so that*

$$K\subset\bigcup_{i=1}^{M}B(x_i,\varepsilon)$$

where $B(x,\varepsilon)=\{y\in K \text{ s.t. } \delta(x,y)\leqslant\varepsilon\}$.

An especially important case is when K is a subset of $\mathbf{R}^n$ and δ is induced by a norm. The next lemma is proved by a volumetric argument (see [13], Lemma 9.5).

Lemma A1. *If $\|\cdot\|$ is a norm on $\mathbf{R}^n$ with unit ball K, then for every $\varepsilon>0$, $N(K,\|\cdot\|,\varepsilon)\leqslant(1+2/\varepsilon)^n$.*

The following theorem gives upper bounds on Bernoulli averages involving covering numbers. For a proof, see Lemma 4.5 and Theorem 11.17 in [13].

Theorem A2 (Dudley's inequality)**.** *Let (ε_i) be independent Bernoulli random variables and K be a compact subset of $\mathbf{R}^n$. Denote by $(x_1,\dots,x_n)$ the coordinates of a vector $x\in\mathbf{R}^n$. Then for some absolute constant C,*

$$\mathbf{E}\max_{x\in K}\sum_{i=1}^{n}\varepsilon_i x_i\leqslant C\int_0^\infty\sqrt{\log N(K,|\cdot|,\varepsilon)}d\varepsilon$$

where $|\cdot|$ denotes the Euclidean norm on $\mathbf{R}^n$.

4.2. **2-convexity.**

Definition. *A Banach space $(E,\|\cdot\|)$ is said to be* 2-convex *with constant λ if for any $y,z\in E$, we have*

$$\|y\|^2+\lambda^{-2}\|z\|^2\leqslant\frac{1}{2}(\|y+z\|^2+\|y-z\|^2).$$

The smallest such λ is called the 2-convexity constant of E and denoted by $\lambda(E)$.

We say shortly that "E is 2-convex" while the usual terminology should be "E has a modulus of convexity of power type 2". This should not be confused with the notion of 2-convexity for Banach lattices [14].

It follows from the parallelogram identity that a Hilbert space is 2-convex with constant 1. Other examples are ℓ_p and S_p^d for $1<p\leqslant 2$. The next theorem has been proved by Ball, Carlen and Lieb [4], refining on early work by Tomczak–Jaegermann [16].

Theorem A3. *For $p\leqslant 2$, the following inequality holds for $A,B\in\mathcal{M}(\mathbf{C}^d)$*

$$\|A\|_p^2+(p-1)\|B\|_p^2\leqslant\frac{1}{2}\left(\|A+B\|_p^2+\|A-B\|_p^2\right).$$

Therefore, S_p^d is 2-convex with constant $1/\sqrt{p-1}$.

This property nicely dualizes. Indeed, it is easily checked (see [4], lemma 5) that E is 2-convex with constant λ if any only if, for every $y,z\in E^*$,

$$\|y\|_*^2+\lambda^2\|z\|_*^2\geqslant\frac{1}{2}(\|y+z\|_*^2+\|y-z\|_*^2).$$

In this case, E^* is said to be *2-smooth* with constant λ.

4.3. **Type 2.**

Definition. *A Banach space $(E, \|\cdot\|)$ is said to have* type 2 *if there exists a constant T_2 so that for any finite sequence $y_1, \dots, y_N$ of vectors of E, we have*

$$\left(\mathbf{E}\left\|\sum_{i=1}^N \varepsilon_i y_i\right\|^2\right)^{1/2} \leqslant T_2\left(\sum_{i=1}^N \|y_i\|^2\right)^{1/2}. \tag{7}$$

The smallest possible T_2 is called the type 2 constant of E and denoted $T_2(E)$. Here, the expectation $\mathbf{E}$ *is taken with respect to a sequence (ε_i) of independent Bernoulli random variables.*

It follows from the (generalized) parallelogram identity that a Hilbert space has type 2 with constant 1, and there is actually equality in (7). If a Banach space E is 2-convex, then E^* is 2-smooth. It is easily checked (by induction on the number of vectors involved) that a 2-smooth Banach space has type 2 with the same constant. We therefore have the inequality $T_2(E^*) \leqslant \lambda(E)$. In particular, theorem A3 implies the following result, first proved by Tomczak-Jaegermann [16] with a worse constant.

Theorem A4. *If $q \geqslant 2$, then S_q^d has type 2 with the estimate*

$$T_2(S_q^d) \leqslant \sqrt{q-1}.$$

4.4. **An entropy lemma.** The following lemma plays a key role in our proof. It appears as Lemma 1 in [10].

Lemma A5. *Let E be a Banach space with unit ball $B(E)$. Assume that E is 2-convex with constant $\lambda(E)$. Let $x_1, \dots, x_N$ be elements of E^*, and define a norm $|||\cdot|||$ on E as*

$$|||y||| = \max_{1\leqslant i\leqslant N} |x_i(y)|.$$

Then for any $\varepsilon > 0$ we have for some absolute constant C

$$\varepsilon\sqrt{\log N(B(E), |||\cdot|||, \varepsilon)} \leqslant C\lambda(E)^2 T_2(E^*)\sqrt{\log N}\max_{1\leqslant i\leqslant N}\|x_i\|_{E^*}. \tag{8}$$

The proof of lemma A5 is based on a duality argument for covering numbers coming from [6]. A positive answer to the duality conjecture for covering numbers (see [3] for a statement of the conjecture and recent results) would imply that the inequality (8) is valid without the factor $\lambda(E)^2$. This would improve our estimate in theorem 2 to $N \geqslant Cd(\log d)^4/\varepsilon^2$.

References

[1] A. Ambainis, M. Mosca, A. Tapp and R. de Wolf, *Private quantum channels*. 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), 547–553, IEEE Comput. Soc. Press.

[2] A. Ambainis and A. Smith, *Small Pseudo-random Families of Matrices: Derandomizing Approximate Quantum Encryption*, Proceedings of RANDOM'04, 249–260.

[3] S. Artstein, V. Milman, S. Szarek and N. Tomczak–Jaegermann, *On convexified packing and entropy duality*. Geom. Funct. Anal. **14** (2004), no. 5, 1134–1141.

[4] K. Ball, E. Carlen and E. Lieb, *Sharp uniform convexity and smoothness inequalities for trace norms*. Invent. Math. **115** (1994), no. 3, 463–482.

[5] R. Bhatia, *Matrix analysis*. Graduate Texts in Mathematics **169**. Springer-Verlag, 1997.

[6] J. Bourgain, A. Pajor, S. Szarek and N. Tomczak-Jaegermann, *On the duality problem for entropy numbers of operators*. Geometric aspects of functional analysis (1987–88), 50–63, Lecture Notes in Math. **1376** (1989).

[7] M. D. Choi, *Completely positive linear maps on complex matrices*. Linear Algebra and Appl. **10** (1975), 285–290.

[8] P. Dickinson and A. Nayak, *Approximate Randomization of Quantum States With Fewer Bits of Key*, AIP Conference Proceedings **864**, 18–36 (2006).

[9] R. Durrett, *Probability. Theory and examples*, The Wadsworth & Brooks/Cole Statistics/Probability Series, 1991.

[10] O. Guédon, S. Mendelson, A. Pajor and N. Tomczak–Jaegermann, *Majorizing measures and proportional subsets of bounded orthonormal systems*, preprint (2008).
[11] O. Guédon and M. Rudelson, *L_p-moments of random vectors via majorizing measures*, Adv. Math. **208** (2007), no. 2, 798–823.
[12] P. Hayden, D. Leung, P. W. Shor and A. Winter, *Randomizing quantum states: constructions and applications*, Comm. Math. Phys. **250** (2004), 371–391.
[13] M. Ledoux and M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **23**. Springer-Verlag, 1991.
[14] J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces. II. Function spaces*. Ergebnisse der Mathematik und ihrer Grenzgebiete **97**. Springer-Verlag, 1979.
[15] M. Rudelson, *Random vectors in the isotropic position*. J. Funct. Anal. **164** (1999), no. 1, 60–72.
[16] N. Tomczak-Jaegermann, *The moduli of smoothness and convexity and the Rademacher averages of trace classes $S_p(1 \leqslant p < \infty)$*. Studia Math. **50** (1974), 163–182.
[17] A. Winter, *The maximum output p-norm of quantum channels is not multiplicative for any p¿ 2*, arxiv 0707.0402